Hawley, S., Sanni Ali, M., Berencsi, K., Judge, A., & Prieto-Alhambra, D. (2019). Sample size and power considerations for ordinary least squares interrupted time-series analysis: a simulation study. *Journal of Clinical Epidemiology*, *11*, 197-205. https://doi.org/10.2147/CLEP.S176723

Sample size and power considerations for ordinary least squares interrupted time-series analysis: a simulation study

Samuel Hawley [1], M. Sanni Ali [1,2], Klara Berencsi [1], Andrew Judge[1,3,4], Daniel Prieto-Alhambra [1,5]

Affiliations:

1. Centre for Statistics in Medicine, NDORMS, University of Oxford, UK
2. Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, UK
3. MRC Lifecourse Epidemiology Unit, University of Southampton, UK
4. Translational Health Sciences, University of Bristol, UK
5. GREMPAL Research Group, Idiap Jordi Gol and CIBERFes, Universitat Autònoma de Barcelona and Instituto de Salud Carlos III, Barcelona, Spain

Key words:
Epidemiology, interrupted time series, sample size, power, bias

Running header:
sample size considerations in interrupted time series

Corresponding Author:
Samuel Hawley, Botnar Research Centre, Windmill Road, Oxford, OX3 7LD (ORCHID: 0000-0002-7034-6168)

Word count: 3,304
Number of tables: 0
Number of figures: 5
Online supplementary material: yes

Summary/Abstract

Interrupted time-series (ITS) analysis is being increasingly used in epidemiology. Despite its growing popularity, there is a scarcity of guidance on power and sample size considerations within the ITS framework. Our aim here was to assess the statistical power to detect an intervention effect under various real life ITS scenarios. ITS datasets were created using Monte-Carlo simulations to generate cumulative incidence (outcome) values over time. We generated 1,000 datasets per scenario, varying the number of time points, average sample size per time point, average relative reduction post-intervention, location of intervention in the time-series and reduction mediated via a: (i) slope change and (ii) step change. Performance measures included power and % bias. We found that sample size per timepoint had a large impact on power. Even in scenarios with 12 pre-intervention and 12 post-intervention timepoints with moderate intervention effect sizes, most analyses were underpowered if the sample size per timepoint was low. We conclude that various factors need to be collectively considered to ensure adequate power for an ITS study. We have demonstrated a means of providing insight on underlying sample size requirements in ordinary least squares ITS analysis based on pre-specified parameters and have developed Stata code to estimate this.

INTRODUCTION

Interrupted time-series (ITS) analysis is being increasingly used in epidemiology. (1-3) It is an accessible and intuitive method that can be straight-forward to implement and has considerable strengths (4). A common application is when population-level repeated measures of an outcome and/or exposure are available over time, both before and after some well-defined intervention such as a health policy change (1, 2, 5) or a naturally occurring event of interest. (6, 7)

Despite the substantial growth in use of ITS methods, relatively little practical guidance has been developed in terms of methodological standards within the ITS framework, (1, 3) including a scarcity of guidance on required sample size. Sample size planning is often a key component of designing a study and should be conducted prior to analysis (8), although this is an aspect very often overlooked in ITS studies, with many being underpowered (9).

Information on the power associated with various numbers of repeated measures of an outcome (i.e. timepoints) has been previously reported (10), with rules of thumb concerning the minimum number of pre- and post-intervention timepoints needed, such as three, (3) six, (11) eight, (12) and $\geq$10. (9) However, researchers seeking to aggregate patient-level data into a population-level timeseries in order to conduct an ITS are confronted with the practical issue of considering a suitable underlying sample size of subjects/patients per aggregate timepoint (13). While longer time-series have

been shown to have more power than short time-series, it seems reasonable to propose that ITS analyses (even those with many timepoints) with only a small number of subjects per timepoint may contain so much noise as to render it improbable of detecting a true impact of an intervention under study. While the ITS method has many strengths, if a given analysis is not adequately powered it may lead to publication of weak and spurious findings (14, 15).

Given this paucity of guidance on sample size calculation, our aim here was to use a simulation approach to estimate power in an ITS analysis case-study of repeated measures of cumulative incidence generated from routinely collected healthcare data. We aimed to quantify the power available in relation to the underlying sample size per timepoint, while varying a number of other key parameters of interest. Furthermore, we set out to make available Stata code to be readily usable by epidemiologists as a tool to generate estimates of required sample size for similar ITS applications.

METHODS

Study design

We used Monte-Carlo simulations, the strengths of which have been well described previously (16, 17). Briefly, simulation studies involve generating data with known characteristics defined by pre-specified input parameter values. Consequently, because the truth regarding these characteristics is known, it is possible to empirically

evaluate the performance of a given statistical model when fitted to the simulated data (18, 19).

Aims

Our aim was to describe the power associated with the mean sample size per timepoint to detect a change in: (i) level and (ii) trend in an outcome (cumulative incidence) following a defined intervention in the ITS framework, using OLS regression. We considered a range of values for various other factors such as total number of timepoints, effect size and location of intervention in the time-series. We set out to apply the methods within the context of a specific case-study using a recent ITS analysis where we evaluated the impact of a UK NICE technology appraisal on the cumulative incidence of joint replacement within the Clinical Practice Research Datalink (CPRD). (20)

ITS scenarios

There are many factors within an ordinary least squares ITS framework that could conceivably influence the power to detect the impact of an intervention. Although the following is not an exhaustive list, we here describe the main factors that we investigated.

1. Total number of timepoints in the time-series, N (figure 1A & 1B)

As described in the introduction, the ITS approach relies on repeated observations of an outcome event over time, usually at equally spaced intervals such as days, weeks, months, quarters, years, etc. We investigated nine values for total number of timepoints (N), ranging from 6 to 50.

2. Number of subjects per timepoint, n (figure 1C & 1D)

The sample size per timepoint will impact the accuracy of outcome estimates and hence the dispersion of a given time-series. It is therefore an important factor influencing the power to detect an 'interruption'. We investigated 11 values for n, ranging from approximately 150 to 5700 patients per timepoint, which for our specific case-study corresponded to a mean number of outcome events per timepoint that ranged from 5 to 200 (supplementary file 1).

3. Nature of intervention impact (figure 1A & 1B versus 1C & 1D)

The impact of an intervention can be modelled as a 'step' change in the level of outcome and/or a 'slope' change in the trend of outcome. (4, 21) More complex realities can be incorporated such as multiple interventions, waning or delayed effects and non-linear responses. (2, 21) However, for the purposes of the current work we only considered intervention effects mediated through either: (i) a step change or (ii) a slope change

4. Effect size – i.e. magnitude of intervention impact

One of the assumptions of ITS analysis is that the pre-intervention level and trend of outcome can be used to predict post-intervention counterfactual estimates, i.e. what outcomes would be expected in the post-intervention period had the intervention not occurred. (2, 21) The impact of intervention can then be expressed as the difference between the estimated counterfactual outcome value for a given post-intervention timepoint versus the estimated modelled outcome value for the same timepoint using the observed data. (22) In practice this has often been done for the mid-point of the post-intervention period in order to yield an average post-intervention change. (5, 20, 23) We therefore used the magnitude of this average post-intervention change expressed as a relative % to express effect size, defined for mid-time-series interventions as the step or slope change resulting in a -15%, -34%, -50% and -75% reduction.

5. Mean pre-intervention level and trend of outcome

The absolute pre-intervention level of outcome is an important factor. For example, a relative 50% reduction of a common outcome should be easier to detect than a relative 50% reduction of a rare outcome. Furthermore, a pre-intervention trend in outcome may exist, which may also have an effect on power. We therefore considered two parameters: the mean pre-intervention outcome value (defined using the pre-intervention mid-point) in conjunction with a pre-intervention trend parameter. In

main analyses we here only explored scenarios (based on our prior CPRD study (20)) where mean pre-intervention cumulative incidence was 3.5% and where there was either: (i) no pre-intervention trend (for step change scenarios) or (ii) an upward trend (for slope change scenarios) (figure 1). We scaled trend parameters according to N so that absolute pre-intervention values were constant across all mid-time-series intervention scenarios. Exact parameter values for these are provided in supplementary file 1.

6. Location of intervention in time-series

Related to N, location of intervention in the time-series may also have an impact on power as this will affect the balance in number of pre-intervention and post-intervention timepoints to be modelled. Locations investigated were at: one-third, mid-way and two-thirds from the beginning of the time-series. For trend change scenarios in our case-study, we used the same pre-intervention and post-intervention trends when investigating early/late interventions as per the corresponding mid-way intervention setting within each N scenario (supplementary file 1).

Data Generating Process

Data were generated in Stata v15.2, the general principles of which have been described elsewhere (24). Empty time-series datasets were created of length N (total number of timepoints). Three ITS variables were inserted: timepoint identifier (integer), post-intervention indicator (binary) and post-intervention timepoint

identifier (integer) (21). The timepoint identifier was created first, then used in combination with the 'location of intervention' parameter to generate the other two ITS variables. The underlying sample size for each timepoint (n_t) was simulated from a normal distribution with mean n (a key parameter of interest; 11 values investigated) and standard deviation of n/3. The number of outcome events occurring per timepoint were drawn as a binomial random variate (n_t, p_t), where n_t represents the sample size and p_t the probability of outcome. p_t was a linear function defined using the ITS variables in combination with other scenario-specific parameter values (equation included in supplementary file 1). The number of events per timepoint and n_t were used to derive the cumulative incidence time-series. A total of 1,000 Monte-Carlo repetitions were carried out for each unique scenario.

Methods of analysis

A segmented linear regression model was fitted to each created dataset. This took the form of model (1) for step change scenarios and model (2) for slope change scenarios:

(1) $Y_t = \beta_0 + \beta_1 * \text{timepoint}_t + \beta_2 * \text{intervention\_indicator}_t + e_t$.

(2) $Y_t = \beta_0 + \beta_1 * \text{timepoint}_t + \beta_3 * \text{post\_intervention\_timepoint}_t + e_t$.

Here, Y_t is the value of outcome at timepoint t. β_0 estimates the level of the outcome just before the beginning of the time-series. β_1 estimates the pre-intervention trend, β_2 the change in level between the time point immediately before vs. after the

intervention and β_3 the change in trend occurring immediately after the intervention. e_t is the error term.

Estimands

The target of inference was change in outcome following an intervention, specifically testing the null hypothesis of no change (i.e. β_2=zero [model 1] or β_3=zero [model 2]). The outcome at each timepoint was a proportion, which in our case-study was the 5-year cumulative incidence of joint replacement within rheumatoid arthritis patients. (20)

Performance

The coefficients, standard error and p-values from these models were stored and the empirical power to reject the null hypothesis of no post-intervention change was calculated as the proportion of simulations where the *P*-value for the intervention variable coefficient (step/slope change) was <0.05. (19, 24, 25) This was represented graphically as contour plots across scenarios according to N and n. We truncated the Y axis (depicting sample size) of main graphs although the full axis was used for graphs in the supplementary material. For convenience of comparison, additional presentation was made for power according to different effect size and location scenarios while keeping N constant (N=28). Also calculated for midway step and slope change scenarios (while keeping N constant) was the percentage bias (19) of the regression coefficients, defined as:

$$\% \text{ Bias} = \left[\frac{\text{average estimate across simulations} - \text{true parameter value}}{\text{true parameter value}} \right] * 100$$

Sensitivity analysis

In order to explore the impact of pre-intervention level of outcome, we repeated main analyses investigating power for slope and step changes while keeping N constant (N=28) but varying pre-intervention level from 3.5% to (i) 8% and (ii) 20%. These main analyses were also repeated where the intervention impact was in the form of an increase in outcome rather than reduction.

Stata programme

Although we based the present analyses on a case-study exploring a range of parameter values adapted from our prior CPRD study as specified above, we also developed a Stata programme (supplementary file 6) with associated documentation (supplementary file 5) in order to provide a ready-to-use means for assessing power associated with any valid list of (nine) input parameter values, as described in the supplementary file 5.

RESULTS

Results from our case-study are presented below describing the impact of N and n on power within several ITS scenarios. Although the main results pertain to a setting where the mean pre-intervention level of outcome for mid-time-series interventions

was 3.5%, the Stata programme developed can be used to explore alternative input parameter values (supplementary file 6).

Slope change

As expected, power increased as N and/or n increased (figure 2A) and as effect sizes became larger (figure 3A). Results for different N and n combinations for each effect size investigated are provided in supplementary file 2. These indicated that nearly all mid-time-series intervention scenarios with a large effect size (-75%) had at least 80% power when there were >24 total timepoints, even when there was a very small sample size per timepoint (~150 subjects, corresponding here to only 5 outcome events per timepoint). However, when the effect size was small (-15%) then to achieve 80% power an analysis had to either contain a large N or very large n (supplementary file 2). While keeping other factors constant (effect size =-34% and N=28), power was greater in scenarios with mid-time-series interventions, with comparably less power in scenarios with earlier/later interventions (figure 4A). The % bias in model coefficients was small and this trended towards zero as sample size increased (Figure 5)

Step change

Similar to slope change scenarios, power increased as N and n became larger (figure 2B) or as the effect size was larger (figure 3B). Generally, there was less power in step change scenarios than in corresponding slope change scenarios (figure 2A versus 2B), with nearly all mid-time-series intervention scenarios being inadequately powered

when the effect-size was only -15% (figure 3B, supplementary file 3). Even when effect sizes were large and number of timepoints was moderate (14 pre-intervention and 14 post- intervention timepoints), analyses were underpowered if sample size per timepoint was low (figure 3B, supplementary figure 3). Interestingly, little difference was found in power following an early or late intervention as compared to when the intervention occurred mid-way through (figure 4). The % bias in model coefficients was small and this trended towards zero as sample size increased (Figure 5)

DISCUSSION

Main findings

This study demonstrates that simple rules regarding the number of timepoints are not adequate by themselves to denote an ITS analysis as sufficiently powered. Other factors such as the sample size per timepoint, expected effect size, location of intervention in the time-series and pre-intervention trends need to be considered. For example, in our case-study where mean pre-intervention level of outcome was 3.5%, to achieve 80% power to detect a relative 34% post-intervention step change reduction, with 14 pre- and 14 post-intervention timepoints, one needed over 1,000 subjects per timepoint (i.e. >28,000 total subjects), which may or may not be realistic for a given study. However, three pre- and post-intervention timepoints were equally sufficient to achieve 80% power in relatively rare situations of large intervention effect sizes combined with very large sample sizes per timepoint (supplementary files 2-4).

These results underline the importance of robust pre-study sample size planning. Estimates arising from scenarios with a very small n were only very slightly biased, which disappeared as n increased (figure 3).

That power increases as N increases is an expected finding and has previously been shown for fixed ratios of effect size to the standard deviation of the timeseries (10, 26). However, we've here addressed the previously unknown trade-off between N and n. This is an important consideration and a helpful development. Firstly, because the standard deviation of a given number of population-level timepoints may likely be difficult for applied researchers to estimate in advance of a proposed ITS study. Secondly, because exploring this trade-off between N and n informs to what extent it may be beneficial (in terms of power) when generating an aggregate ITS dataset to sacrifice sample size per timepoint in order to increase the number of timepoints (or vice-versa). It allows a combination of N and n to be selected to optimise power. Although the exact nuances of this unique trade-off were scenario specific, in most cases only very little gain in power was achieved when a time-series was lengthened at the expense of timepoint sample size, although gains were more noticeable where a very short time-series was lengthened.

To our knowledge, a differential power according to whether an intervention impact is mediated via a slope or step change has not previously been investigated. We found power was greater in slope change scenarios, a likely explanation being that our effect size was the average difference between post-intervention values and counterfactuals,

which in the case of slope change scenarios continued to increase as per the pre-intervention slope and therefore made detection of a change more probable.

Within scenarios with a slope change we found power to be greater in settings with a balanced number of pre-intervention and post-intervention timepoints (as opposed to earlier/later interventions), while the location of the intervention had little impact on power to detect step changes and was even marginally greater when the intervention occurred early. Although this was unexpected, it is not without some support from previous work (10).

Limitations

Our study is subject to various limitations. Each timepoint was a cumulative incidence, and given that individual subjects/patients could only be included in a single timepoint we treated timepoints to be independent. As such we've not explored what impact autocorrelation may have on estimates, although this remains a subject for further investigation. Despite the availability of ITS approaches that explicitly model autocorrelation, such as autoregressive integrated moving average (ARIMA) models (27), it would seem that where the assumptions of OLS regression are met then this is preferable for epidemiological studies where the goal is likely to be causal inference rather than future prediction. Indeed, while autocorrelation needs to be addressed where present, it has been noted that in epidemiological studies it can often be accounted for by controlling for other variables (2), and interestingly of a recent review of over 200 drug utilization studies implementing ITS analysis, 50% were found to use

segmented linear regression (1). Specification of ARIMA models are frequently cited to require a minimum of 50 timepoints (28), with >100 being preferable (27), yet it is common to have less than this minimum available in epidemiology contexts using routinely collected data (10, 21, 23, 29). For these reasons, our focus here was on 'short' time-series where we've considered 50 timepoints as a maximum and used Durbin Watson statistics to confirm first-order autocorrelation was not present. Previous work has investigated the relationship between the number of timepoints and power in the presence of autocorrelation (10, 30), where positive autocorrelation has been shown to reduce power and negative autocorrelation to increase power (10). Similarly, we've not considered seasonality nor situations where there may be a delay or waning intervention effect.

Another limitation is that our definition of effect size as the difference between post-intervention timepoints and counterfactual timepoints (i.e. what would have been observed had pre-intervention level/slope continued uninterrupted) involves extrapolation and therefore uncertainty. While this is often done in practice, with uncertainty of model estimates expressed using confidence intervals, (22) there is still the assumption that pre-intervention trends would have continued unchanged.

We've also only investigated scenarios where the repeated outcome measure is a cumulative incidence (i.e. a proportion). This is a very common epidemiological measure, but incorporating other common measures such as person-year rates, means

(for example length of hospital stay or drug doses prescribed) and frequencies is a logical next step and remains the subject for imminent further investigation.

Strengths

The disentangling of N and n is a key strength and novel aspect of the current study, as is the separate consideration of post-intervention step and slope changes. The development and inclusion of a Stata programme is an important feature of the investigation, facilitating researchers to estimate sample size requirements for future ITS studies and thereby promoting the avoidance of carrying out underpowered analyses. We are currently working on using this tool as the basis for an online calculator. It is also worth mentioning that we based the parameter values for our case-study on a "real world" clinical scenario (20) in order to increase the applicability of the findings, rather than starting from arbitrary parameter values.

Conclusion

Multiple factors influence the power of OLS ITS analysis and these should be collectively taken into account when considering the feasibility of a proposed ITS study. We have demonstrated how a simulation approach can be used to estimate the power available within specific ITS scenarios and have developed Stata code to facilitate pre-analysis sample size planning of future ITS studies within similar applications.

Funding

This project was partially supported by Oxford NIHR Biomedical Research Unit. Andrew Judge was partially supported by the NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

Competing interests

SH, SA and KB have no conflicts of interest. AJ has received consultancy fees from Freshfields Bruckhaus Deringer, and is a member of the Data Safety and Monitoring Board (which involved receipt of fees) from Anthera Pharmaceuticals, INC. outside the submitted work. DPAs research group has received unrestricted research grants from Servier Laboratoires, AMGEN and UCB Pharma.

Author contributions:

Study design: SH, DPA, AJ; Data simulation: SH, SA, KB; Data analysis: SH, KB, SA; Drafting manuscript: SH; Correcting and/or approving final manuscript: All authors

References

1. Jandoc R, Burden AM, Mamdani M, Levesque LE, Cadarette SM. Interrupted time series analysis in drug utilization research is increasing: systematic review and recommendations. J Clin Epidemiol. 2015;68(8):950-6.
2. Lopez Bernal J, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. International journal of epidemiology. 2016.
3. Ewusie JE, Blondal E, Soobiah C, Beyene J, Thabane L, Straus SE, et al. Methods, applications, interpretations and challenges of interrupted time series (ITS) data: protocol for a scoping review. BMJ Open. 2017;7(6):e016018.
4. Kontopantelis E, Doran T, Springate DA, Buchan I, Reeves D. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. Bmj. 2015;350:h2750.
5. Hawton K, Bergen H, Simkin S, Dodd S, Pocock P, Bernal W, et al. Long term effect of reduced pack sizes of paracetamol on poisoning deaths and liver transplant activity in England and Wales: interrupted time series analyses. Bmj. 2013;346:f403.
6. Laliotis I, Ioannidis JPA, Stavropoulou C. Total and cause-specific mortality before and after the onset of the Greek economic crisis: an interrupted time-series analysis. Lancet Public Health. 2016;1(2):e56-e65.
7. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. J Epidemiol Community Health. 2012;66(12):1182-6.
8. Lenth RV. Some practical guidelines for effective sample size determination. Am Stat. 2001;55(3):187-93.
9. Ramsay CR, Matowe L, Grilli R, Grimshaw JM, Thomas RE. Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behavior change strategies. Int J Technol Assess Health Care. 2003;19(4):613-23.
10. Zhang F, Wagner AK, Ross-Degnan D. Simulation-based power calculation for designing interrupted time series analyses of health policy interventions. J Clin Epidemiol. 2011;64(11):1252-61.
11. Fretheim A, Zhang F, Ross-Degnan D, Oxman AD, Cheyne H, Foy R, et al. A reanalysis of cluster randomized trials showed interrupted time-series studies were valuable in health system evaluation. J Clin Epidemiol. 2015;68(3):324-33.
12. Penfold RB, Zhang F. Use of Interrupted Time Series Analysis in Evaluating Health Care Quality Improvements. Acad Pediatr. 2013;13(6):S38-S44.
13. Cordtz RL, Hawley S, Prieto-Alhambra D, Hojgaard P, Zobbe K, Overgaard S, et al. Incidence of hip and knee replacement in patients with rheumatoid arthritis following the introduction of biological DMARDs: an interrupted time-series analysis using nationwide Danish healthcare registers. Ann Rheum Dis. 2018;77(5):684-9.

14. Van Calster B, Steyerberg EW, Collins GS, Smits T. Consequences of relying on statistical significance: Some illustrations. Eur J Clin Invest. 2018;48(5):e12912.
15. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013;14(5):365-76.
16. Chang M. Monte Carlo Simulation for the Pharmaceutical Industry: Concepts, Algorithms, and Case Studies: CRC Press; 2010.
17. Ripley BD. Stochastic Simulation. New York, United States: John Wiley and Sons Ltd; 2006.
18. Morris T, White I, Crowther M. Using simulation studies to evaluate statistical methods. Tutorial in Biostatistics. 2017.
19. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. Statistics in medicine. 2006;25(24):4279-92.
20. Hawley S, Cordtz R, Lene D, Edwards CJ, Arden NK, Delmestri A, et al. Association between NICE guidance on biologic therapies with rates of hip and knee replacement among rheumatoid arthritis patients in England and Wales: An interrupted time-series analysis. American College of Rheumatology; Washington DC: Arthritis Rheumatol.; 2016.
21. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. J Clin Pharm Ther. 2002;27(4):299-309.
22. Zhang F, Wagner AK, Soumerai SB, Ross-Degnan D. Methods for estimating confidence intervals in interrupted time series analyses of health interventions. J Clin Epidemiol. 2009;62(2):143-8.
23. Hawton K, Bergen H, Simkin S, Brock A, Griffiths C, Romeri E, et al. Effect of withdrawal of co-proxamol on prescribing and deaths from drug poisoning in England and Wales: time series analysis. Bmj. 2009;338:b2270.
24. Feiveson AH. Power by simulation. The Stata Journal. 2002;2(2):107-24.
25. Sayers A, Crowther MJ, Judge A, Whitehouse MR, Blom AW. Determining the sample size required to establish whether a medical device is non-inferior to an external benchmark. BMJ Open. 2017;7(8):e015397.
26. McLeod AI, Vingilis ER. Power computations in time series analyses for traffic safety interventions. Accid Anal Prev. 2008;40(3):1244-8.
27. Box GEPJ, G.M.; Reinsel, G.C.; Ljung, G.M. Time Series Analysis: Forecasting and Control. 5th ed: WILEY; 2015 29/06/2015.
28. Chatfield C. The Analysis of Time Series: an introduction. Chatfield CT, M.; Zidek, J., editor. Taylor & Francis e-Library: CHAPMAN & HALL/CRC; 2009.
29. Hawley S, Leal J, Delmestri A, Prieto-Alhambra D, Arden NK, Cooper C, et al. Anti-Osteoporosis Medication Prescriptions and Incidence of Subsequent Fracture Among Primary Hip Fracture Patients in England and Wales: An Interrupted Time-Series Analysis. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2016;31(11):2008-15.
30. McLeod AI, Vingilis ER. Power computations for intervention analysis. Technometrics. 2005;47(2):174-81.